An Invitation to Statistics in Wasserstein Space

Shen Ziliang 2021213218

Shanghai University of Finance and Economics

March 20, 2022

Outline

Content

The Fréchet Functional

Definition 3.1.1 (Empirical Fréchet Functional and Mean)

The Fréchet functional associated with measures $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ is

$$F : \mathcal{W}_2(\mathcal{X}) \to \mathbb{R} \qquad F(\gamma) = \frac{1}{2N} \sum_{i=1}^{N} W_2^2(\gamma, \mu^i), \qquad \gamma \in \mathcal{W}_2(\mathcal{X}). (3.1)$$

A Fréchet mean of $\mu^1, \dots, \mu^N$ is a minimiser of F in $\mathcal{W}_2(\mathcal{X})$ (if it exists).

In analysis, a Fréchet mean is often called a barycentre. We shall use the terminology of "Fréchet mean" that is arguably more popular in statistics. The factor $1(2N)$ is irrelevant for the definition of Fréchet mean. It is introduced in order to have simpler expressions for the derivatives (Theorems 3.1.14 and 3.2.13) and to be compatible with a population version $\mathbb{E} W_2^2(\gamma, \Lambda)/2$(population versions).

The Fréchet Functional

The first reference that deals with empirical Fréchet means in $\mathcal{W}_2(\mathbb{R}^d)$ is the seminal paper of Agueh and Carlier. They treat the more general weighted Fréchet functional

$$F(\gamma) = \frac{1}{2}\sum_{i=1}^{N} w_i W_2^2(\gamma, \mu^i), \qquad 0 \le w_i, \quad \sum_{i=1}^{N} w_i = 1$$

but, for simplicity, we shall focus on the case of equal weights. (If all the w_i's are rational, then the weighted functional can be encompassed in (3.1) by taking some of the μ^i's to be the same. The case of irrational w_i's is then treated with continuity arguments. Moreover, population versions encapsulates (3.1) as well as the weighted version when Γ can take finitely many values.)

Multimarginal Formulation

multimarginal Monge–Kantorovich problem

Let $\mu^1, \ldots, \mu^N$ be N measures in $\mathcal{W}_2(\mathcal{X})$ and let $\Pi(\mu^1, \ldots, \mu^N)$ be the set of probability measures in $\mathcal{X}^N$ having $\{\mu^i\}_{i=1}^N$ as marginals. The problem is to minimise

$$G(\pi) = \frac{1}{2N^2} \int_{\mathcal{X}^N} \sum_{i<j} \|x_i - x_j\|^2 \, \mathrm{d}\pi(x_1, \ldots, x_N), \qquad \text{over} \quad \pi \in \Pi(\mu^1, \ldots, \mu^N)$$

If $N = 2$, we obtain the Kantorovich problem with quadratic cost. The probabilistic interpretation is that one is given random variables $X_1, , X_N$ with marginal probability laws $\mu^1, \ldots, \mu^N$ and seeks to construct a random vector $Y = (Y_1, \ldots, Y_N)$ on $\mathcal{X}^N$ such that $X_i \stackrel{d}{=} Y_i$ and

$$\frac{1}{2N^2} \mathbb{E} \sum_{i<j} \|Y_i - Y_j\|^2 \leq \frac{1}{2N^2} \mathbb{E} \sum_{i<j} \|Z_i - Z_j\|^2.$$

for any other random vector $Z = (Z_1, \ldots, Z_N)$ such that $X_i \stackrel{d}{=} Z_i$.

Multimarginal Formulation

Intuitively, we seek a random vector with prescribed marginals but maximally correlated entries.
We refer to elements of $\Pi(\mu^1, \ldots, \mu^N)$ (equivalently, joint laws of $X_1, \ldots, X_N$) as multicouplings (of $\mu^1, \ldots, \mu^N$). Just like in the Kantorovich problem, there always exists an optimal multicoupling π. Let us now show how the multimarginal problem is equivalent to the problem of finding the Fréchet mean of $\mu^1, \ldots, \mu^N$. The first thing to observe is that the objective function can be written as

$$G(\pi) = \int_{\mathcal{X}^N} \frac{1}{2N} \sum_{i=1}^{N} \|x_i - M(x)\|^2 \, \mathrm{d}\pi(x), M(x) = M(x_1, \ldots, x_n) = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

Multimarginal Formulation

Proposition 3.1.2 (Fréchet Means and Multicouplings)

Let $\mu^1, \dots, \mu^N \in \mathcal{W}(\mathcal{X})$. Then μ is a Fréchet mean of $\mu^1, \dots, \mu^N$ if and only if there exists an optimal multicoupling $\pi \in \mathcal{W}(\mathcal{X}^N)$ of $\mu^1, \dots, \mu^N$ such that $\mu = M\#\pi$, and furthermore $F(\mu) = G(\pi)$.

Corollary 3.1.3 (Fréchet Means and Moments)

Any finite collection of measures $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ admits a Fréchet mean μ, for all $p \geq 1$

$$\int_{\mathcal{X}} \|x\|^p \, \mathrm{d}\mu(x) \leq \frac{1}{N} \sum_{i=1}^{N} \int_{\mathcal{X}} \|x\|^p \, \mathrm{d}\mu^i(x)$$

and when $p > 1$ equality holds if and only if $\mu^1 = \cdots = \mu^N$.

Multimarginal Formulation

Corollary 3.1.4

The support of any Fréchet mean is included in the set

$$\frac{\mathrm{supp}\mu^1 + \cdots + \mathrm{supp}\mu^N}{N} = \left\{ \frac{x_1 + \cdots + x_N}{N} : x_i \in \mathrm{supp}\mu^i \right\} \subseteq \mathrm{conv} \left(\bigcup_{i=1}^{N} \mathrm{supp}\mu^i \right)$$

In particular, if all the μ^i's are supported on a common convex set K, then so is any of their Fréchet means.

Continuity

Theorem 3.1.5 (Continuity of Fréchet Means)

Suppose that $W_2(\mu_k^i, \mu^i) \to 0$ for $i = 1, \ldots, N$ and let $\overline{\mu}_k$ denote any Fréchet mean of $(\mu_k^1, \ldots, \mu_k^N)$. Then $(\overline{\mu}_k)$ stays in a compact set of $\mathcal{W}_2(\mathcal{X})$, and any limit point is a Fréchet mean of $\mu_k^1, \ldots, \mu_k^N$.
In particular, if $\mu_k^1, \ldots, \mu_k^N$ have a unique Fréchet mean $\overline{\mu}$, then $\overline{\mu}_k \to \overline{\mu}$ in $\mathcal{W}(\mathcal{X})$.

Uniqueness and Regularity

A general situation in which Fréchet means are unique is when the Fréchet functional is strictly convex. In the Wasserstein space, this requires some regularity, but weak convexity holds in general. Absolutely continuous measures on infinite-dimensional $\mathcal{X}$ are defined in Definition 1.6.4.

Proposition 3.1.6 (Convexity of the Fréchet Functional)

Let $\Lambda, \gamma_i \in \mathcal{W}_2(\mathcal{X})$ and $t \in [0,1]$ then

$$W_2^2(t\gamma_1 + (1-t)\gamma_2, \Lambda) \leq tW_2^2(\gamma_1, \Lambda) + (1-t)W_2^2(\gamma_2, \Lambda).$$

When Λ is absolutely continuous, the inequality is strict unless $t \in 0, 1$ or $\gamma_1 = \gamma_2$.

Remark 3.1.7

The Wasserstein distance is not convex along geodesics. That is, if we replace the linear interpolant $t\gamma_1 + (1-t)\gamma_2$ by McCann's interpolant, then $t \mapsto W_2^2(\gamma_t, \Lambda)$ is not necessarily convex .

Uniqueness and Regularity

Proposition 3.1.8 (L_∞-Regularity of Fréchet Means)

Let $\mu^1, \ldots, \mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ and suppose that μ^1 is absolutely continuous with density bounded by M. Then the Fréchet mean of $\{\mu^i\}$ is absolutely continuous with density bounded by $N^d M$ and is consequently a Karcher mean.

The One-Dimensional and the Compatible Case

When $\mathcal{X} = \mathbb{R}$, there is a simple expression for the Fréchet mean because $\mathcal{W}_2(\mathbb{R})$ can be imbedded in a Hilbert space. Indeed, recall that

$$W_2(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|_{L_2(0,1)}$$

In view of that, $\mathcal{W}_2(\mathbb{R})$ can be seen as the convex closed subset of $L_2(0,1)$ formed by equivalence classes of left-continuous nondecreasing functions on (0, 1): any quantile function is left-continuous and nondecreasing, and any such function G can be seen to be the inverse function of the distribution function, the right-continuous inverse of G

$$F(x) = \inf\{t \in (0,1) : G(t) > x\} = \sup\{t \in (0,1) : G(t) \leq x\}$$

Therefore, the Fréchet mean of $\mu^1, \ldots, \mu^N \in \mathcal{W}_2(\mathbb{R})$ is the measure having quantile function

$$F_\mu^{-1} = \frac{1}{N}\sum_{i=1}^{N} F_{\mu^i}$$

The One-Dimensional and the Compatible Case

The Fréchet mean is thus unique. This is no longer true in higher dimension, unless some regularity is imposed on the measures. Boissard et all. noticed that compatibility of $\mu^1, \ldots, \mu^N$ according to Definition2.3.1 allows for a simple solution to the Fréchet mean problem, as in the one-dimensional case. Recall from Proposition 3.1.2 that this is equivalent to the multimarginal problem. Returning to the original form of G, we obtain an easy lower bound for any $\pi \in \Pi(\mu^1, \ldots, \mu^N)$:

$$G(\pi) = \frac{1}{2N^2} \int_{\mathcal{X}^N} \sum_{i<j} \|x_i - x_j\|^2 \, \mathrm{d}\pi(x_1, \ldots, x_N) \geq \frac{1}{2N^2} \sum_{i<j} W_2^2(\mu^i, \mu^j)$$

because the (i,j)-th marginal of π is a coupling of μ^i and μ^j. Thus, if equality above holds for π, then π is optimal and $M\#\pi$ is the Fréchet mean by Proposition 3.1.2. This is indeed the case for $\pi = (\mathbf{i}, \mathbf{t}_{\mu^1}^{\mu^2}, \ldots, \mathbf{t}_{\mu^1}^{\mu^N})\#\mu^1$

because the compatibility gives:

$$\int_{\mathcal{X}^N} \|x_i - x_j\|^2 \, \mathrm{d}\pi(x_1, \ldots, x_N) = \int_{\mathcal{X}} \left\| \mathbf{t}_{\mu^1}^{\mu^i} - \mathbf{t}_{\mu^1}^{\mu^j} \right\|^2 \mathrm{d}\mu^1$$

$$= \int_{\mathcal{X}} \left\| \mathbf{t}_{\mu^1}^{\mu^i} \circ \mathbf{t}_{\mu^j}^{\mu^i} - \mathbf{i} \right\| \mathrm{d}\mu^j = W_2^2(\mu^i, \mu^j).$$

Theorem 3.1.9 (Fréchet Mean of Compatible Measures)

Suppose that $\{\gamma, \mu^1, \ldots, \mu^N\}$ are compatible measures. Then

$$\left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{t}_{\gamma}^{\mu^i} \right] \# \gamma$$

is the Fréchet mean of $(\mu^1, \ldots, \mu^N)$.

The Agueh–Carlier Characterisation

Agueh and Carlier provide a useful sufficient condition for γ to be the Fréchet mean. When $\mathcal{X} = \mathbb{R}^d$, this condition is also necessary , hence characterising Fréchet means in $\mathbb{R}^d$. It will allow us to easily deduce some equivariance results for Fréchet means with respect to *independence (Lemma 3.1.11)* and *rotations (3.1.12)*. More importantly, it provides a sufficient condition under which a local minimum of F is a global minimum (Theorem 3.1.15) and the same idea can be used to relate the population Fréchet mean to the expected value of the optimal maps (Theorem 4.2.4). Recall that ϕ^* denotes the Legendre transform of ϕ, as defined on page 14.

Proposition 3.1.10 (Fréchet Means and Potentials)

Let $\mu^1, \ldots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ be absolutely continuous, let $\gamma \in \mathcal{W}_2(\mathcal{X})$ and denote by ϕ_i^* the convex potentials of $\mathbf{t}_{\mu^i}^{\gamma}$. If $\phi_i = \phi_i^{**}$ are such that

$$\frac{1}{N}\sum_{i=1}^{N} \phi_i(x) \le \frac{1}{2}\|x\|^2, \qquad \forall x \in \mathcal{X}, \qquad \text{withequality}\gamma\text{−almostsurely}$$

The Agueh–Carlier Characterisation

γ is the unique Fréchet mean of $\mu^1, \ldots, \mu^N$.

Lemma 3.1.11 (Independent Fréchet Means)

Let $\mu^1, \ldots, \mu^N$ and $\nu^1, \ldots, \nu^N$ be absolutely continuous measures in $\mathcal{W}_2(\mathbb{R}^{d_1})$ and $\mathcal{W}_2(\mathbb{R}^{d_2})$ with Fréchet means μ and ν, respectively. Then the independent coupling $\mu \otimes \nu$is the Fréchet mean of $\mu^1 \otimes \nu^1, \ldots, \mu^N \otimes \nu^N$ By induction (or a straightforward modification of the proof), one can show that the Fréchet mean of $(\mu^i \otimes \nu^i \otimes \rho^i)$ is $\mu \otimes \nu \otimes \rho$ and so on.

Lemma 3.1.12 (Rotated Fréchet Means)

If μ is the Fréchet mean of the absolutely continuous measures $\mu^1, \ldots, \mu^N$ and U is orthogonal, then $U\#\mu$ is the Fréchet mean of $U\#\mu^1, \ldots, U\#\mu^N$.

Differentiability of the Fréchet Functional and Karcher Means

Definition 3.1.13 (Karcher Mean)

Let F be a Fréchet functional associated with some random measure Λ in $\mathcal{W}_2(\mathcal{X})$ .Then γ is a Karcher mean for Λ if F is differentiable at γ and $F'(\gamma) = 0$.

It turns out that the tangent bundle structure described in Sect2.3 gives rise to a differentiable structure in the Wasserstein space. Fix $\mu^0 \in \mathcal{W}_2(\mathcal{X})$ and consider the function

$$F_0 : \mathcal{W}_2(\mathcal{X}) \to \mathbb{R}, \qquad F_0(\gamma) = \frac{1}{2} W_2^2(\gamma, \mu^0).$$

when   is absolutely continuous,

$$\lim_{W_2(\nu,\gamma)\to 0} \frac{F_0(\nu) - F_0(\gamma) + \int_{\mathcal{X}} \left\langle \mathbf{t}_\gamma^{\mu^0}(x) - x, \mathbf{t}_\gamma^{\nu}(x) - x \right\rangle \mathrm{d}\gamma(x)}{W_2(\nu, \gamma)} = 0.$$

Differentiability of the Fréchet Functional and Karcher Means

The integral above can be seen as the inner product

$$\left\langle \mathbf{t}_{\gamma}^{\mu^0} - \mathbf{i}, \mathbf{t}_{\gamma}^{\nu} - \mathbf{i} \right\rangle$$

in the space $L_2(\gamma)$ that includes as a (closed) subspace the tangent space Tan_{γ}. In terms of this inner product and the log map, we can write

$$F_0(\nu) - F_0(\gamma) = -\left\langle \log_{\gamma}(\mu^0), \log_{\gamma}(\nu) \right\rangle + o(W_2(\nu,\gamma)), \qquad \nu \to \gamma \quad \mathrm{in} \mathcal{W}_2$$

so that F_0 is Fréchet-differentiable at γ with derivative

$$F_0'(\gamma) = -\log_{\gamma}(\mu^0) = -\left(\mathbf{t}_{\gamma}^{\mu^0} - \mathbf{i}\right) \in \mathrm{Tan}_{\gamma}$$

By linearity, one immediately obtains:

Theorem 3.1.14 (Gradient of the Fréchet Functional)

Fix a collection of measures $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathcal{X})$ When $\gamma \in \mathcal{W}_2(\mathcal{X})$ is absolutely continuous, the Fréchet functional

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^{N} W_2^2(\gamma, \mu^i), \qquad \gamma \in \mathcal{W}_2(\mathcal{X})$$

is Fréchet-differentiable and

$$F'(\gamma) = -\frac{1}{N} \sum_{i=1}^{N} \log_\gamma(\mu^i) = -\frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{t}_\gamma^{\mu_i} - \mathbf{i}\right).$$

Theorem 3.1.15 (Optimality Criterion for Karcher Means)

Let $U \subseteq \mathbb{R}^d$ be an open convex set and let $\mu^1, \dots, \mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ be probability measures on U with bounded strictly positive densities $g_1, \dots, g_N$. Suppose that an absolutely continuous Karcher mean γ is supported on U with bounded strictly positive density f there. Then γ is the Fréchet mean of $\mu^1, \dots, \mu^N$ if one of the following holds:

1. $U = \mathbb{R}^d$ and the densities $f, g_1, \dots, g_N$ are of class $C^{0,\alpha}$ for some $\alpha > 0$;
2. U is bounded and the densities $f, g_1, \dots, g_N$ are bounded below on U.